ANL/ALCF/ESP-17/7

Argonne NATIONAL LABORATORY

Next-Generation Cosmology Simulations with HACC: Conquering the Baryon Problem

Technical Report for the ALCF Theta Early Science Program

Argonne Leadership Computing Facility

ALCF Early Science Program (ESP) Technical Report

ESP Technical Reports describe the code development, porting, and optimization done in preparing an ESP project's application code(s) for the next generation ALCF computer system. This report is for a project in the Theta ESP, preparing for the ALCF Theta computer system.

About Argonne National Laboratory

Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

DOCUMENT AVAILABILITY

Online Access: U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free at OSTI.GOV (http://www.osti.gov/), a service of the U.S. Dept. of Energy's Office of Scientific and Technical Information

Reports not in digital format may be purchased by the public from the National Technical Information Service (NTIS):

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312
www.ntis.gov
Phone: (800) 553-NTIS (6847) or (703) 605-6000
Fax: (703) 605-6900
Email: **orders@ntis.gov**

Reports not in digital format are available to DOE and DOE contractors from the Office of Scientific and Technical Information (OSTI):

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
www.osti.gov
Phone: (865) 576-8401
Fax: (865) 576-5728
Email: **reports@osti.gov**

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

ANL/ALCF/ESP-17/7

Next-Generation Cosmology Simulations with HACC: Conquering the Baryon Problem

Technical Report for the ALCF Theta Early Science Program

edited by
Timothy J. Williams and Ramesh Balakrishnan

Argonne Leadership Computing Facility

prepared by
JD Emberson, Nicholas Frontiere, Noe Brucy, Katrin Heitmann, Salman Habib, Hal Finkel, Adrian Pope, Vitali Morozov, and Tom Uram

September 2017

This work was supported in part by the Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

Next-Generation Cosmology Simulations with HACC: Conquering the Baryon Problem

JD Emberson[*1], Nicholas Frontiere[†2], Noe Brucy[‡3], Katrin Heitmann[§4], Salman Habib[¶5], Hal Finkel[‖6], Adrian Pope[**7], Vitali Morozov[††8], and Tom Uram[‡‡9]

[1,6,7,8,9] Leadership Computing Facility, Argonne National Laboratory, Argonne, IL
[2] University of Chicago, Chicago, IL
[3]ENS Rennes, University of Rennes 1, Rennes
[4,5] High Energy Physics Division, Argonne National Laboratory, Argonne, IL

1 Introduction

The next generation of cosmological surveys will start taking data in a few years. These surveys will cover the microwave, optical, and X-ray bands. A major challenge for the interpretation of these new observations will be to disentangle astrophysical effects (e.g., galaxy bias and evolution, baryonic feedback processes) from fundamental physics (dark energy, dark matter, neutrino properties, primordial fluctuations). Efficient large-scale cosmology codes that incorporate the effects of baryons and feedback are crucial to make progress in this area. As part of this Early Science Project on Theta we have built new capabilities for the extreme-scale Hardware/Hybrid Accelerated Cosmology Code (HACC) to allow simultaneous modeling of observations in all of these wavebands at high fidelity. A hydrodynamics capability based on a new, significantly improved, SPH algorithm – Conservative Reproducing Kernel Smoothed Particle Hydrodynamics (CRKSPH) – has been incorporated. This important, new HACC enhancement has been an important step for the next generation of supercomputers. In this report we describe results from the gravity-only version of HACC on Theta, including detailed performance studies, and describe in detail the new implementation of CRK-HACC.

[*]jemberson@anl.gov
[†]frontiere@gmail.com
[‡]noe.brucy@ens-rennes.fr
[§]heitmann@anl.gov
[¶]habib@anl.gov
[‖]hfinkel@anl.gov
[**]apope@anl.gov
[††]morozov@anl.gov
[‡‡]turam@anl.gov

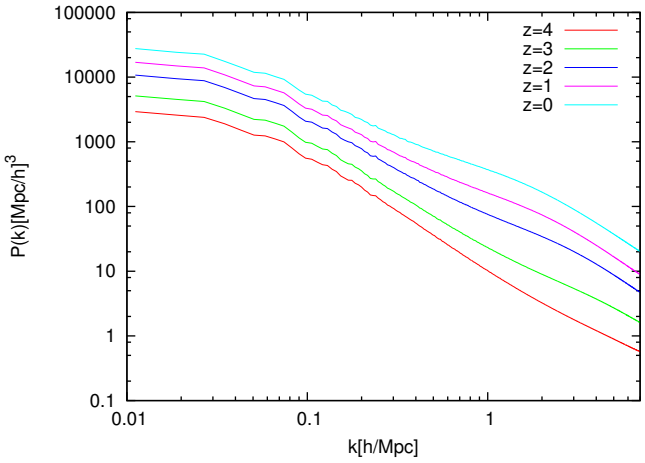

Figure 1: Power spectrum evolution in the BorgCube simulation.
We show the evolution of the nonlinear power spectrum from the gravity-only simulation with HACC on Theta. The time snapshots shown correspond to the same ones that are shown in Fig. 2. At redshift $z = 4$ the onset of nonlinearities on small spatial scales (large k) can be seen around $k \sim 5h\mathrm{Mpc}^{-1}$ which then grows over time. In the project we will investigate how the power spectrum evolves differently when baryonic effects are taken into account.

2 Science Summary

The scientific target of this project is the generation of cosmological maps to study the so-called kinematic Sunyav-Zel'dovich effect. Effects of baryon physics play an important role in this endeavour and a major aim is to study how baryons affect structure formation when taken into account in cosmological simulations. We have set up a simulation called the BorgCube simulation that covers a cosmological volume of $(800h^{-1}\mathrm{Mpc})^3$ and evolves 2304^3 particles on 3072 nodes of Theta. The size of the simulation and the number of nodes was chosen in such a way that we can accommodate a matching hydrodynamic simulation that starts from the same initial conditions. The hydrodynamic simulation is more demanding on memory usage due to the additional particle species and new attributes like temperature and density that have to be considered. In addition, the computing demands increase considerably.

So far, we have finished the gravity-only HACC BorgCube simulation. The analysis was carried out fully on-the-fly. We have stored a large amount of information, from detailed tracking of the structures forming from early to late times to sub-samples of the particle to power spectrum measurements. Figures 1, 2 show two of these results: Figure 1 shows the nonlinear evolution of the matter power spectrum over time while Figure 2 shows the formation of large-scale structures in the simulation. The second figure shows only a small portion of the full simulation volume – the output from one rank – in order to enable the visualization of the high-resolution details of the simulation. The hydrodynamics counterpart is currently being set up and will run in the very near future.

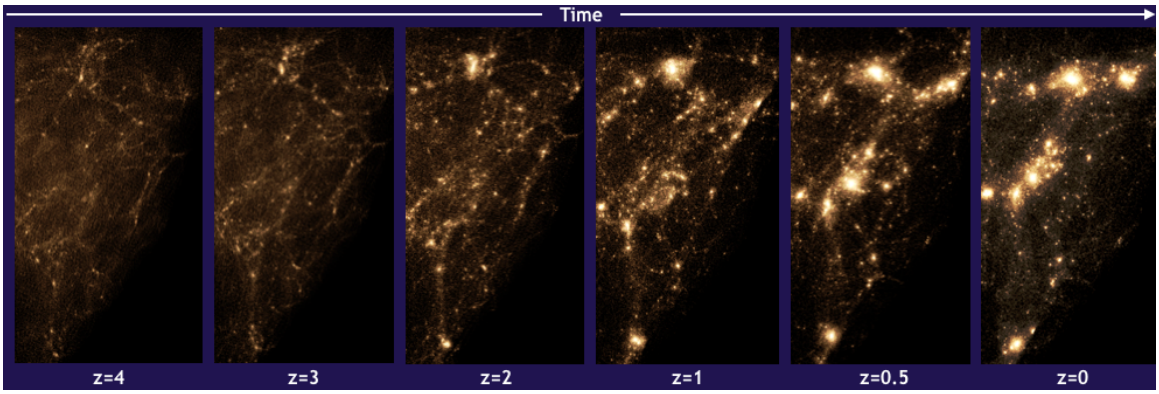

Figure 2: Time evolution of the particles in the BorgCube simulation.
We show a small region of the BorgCube simulation (results from one rank) and how it evolves from early times at redshift $z = 4$ to today at redshift $z = 0$. Structure formation can be clearly seen by following the evolution of a first smooth particle distribution to an ever more clumped distribution. Measuring how baryons effect structure formation is a very important topic in cosmology. Here we show the results for the gravity-only simulation.

3 Codes, Methods and Algorithms

The Hardware/Hybrid Accelerated Cosmology Code (HACC) framework is descended from an approach developed for the heterogeneous architecture of Roadrunner [8], the first computer to break the petaflop barrier. A detailed description of HACC including scaling results on both Titan and Mira is given in Ref. [7]. HACC uses a hybrid parallel algorithmic structure, splitting the force calculation into a specially designed grid-based long/medium range spectral particle-mesh (PM) component that is common to all architectures, and an architecture-specific short-range solver. Modular code design combined with particle caching allows the short-range solvers to be 'hot-swappable' on-node; they are blind to the parallel implementation of the long-range solver. The short-range solvers can use direct particle-particle interactions, i.e., a P^3M algorithm, as on Roadrunner or Titan, or use tree methods as on the IBM BG/Q and Cray XE6 systems (TreePM algorithm). The availability of multiple algorithms within the HACC framework allows us to carry out careful error analyses, for example, the P^3M and the TreePM versions agree to within 0.1% for the nonlinear power spectrum test in the code comparison suite of Ref. [10].

A new hydrodynamic capability has been implemented into HACC. We have developed and tested a new algorithm called Conservative Reproducing Kernel Smoothed Particle Hydrodyanmics (CRKSPH) which addresses some of the short-comings of traditional SPH methods compared to AMR methods. The CRKSPH scheme shows significant improvement in hydrodynamics problems involving shearing and mixing, in addition to shocks. At the same time, it provides the scalability advantage that SPH methods have compared to AMR methods. In the HACC implementation, the gravity solver is unchanged and the new solver can take full advantage of the overload strategy employed by HACC for the baryons. The new implementation also builds on the already available RCB tree implementation. We provide details on this new development in Section 4.

In order to utilize HACC's powerful simulation capabilities, it is essential to have a set of equally powerful analysis tools in place. We have developed a large analysis suite that can be run 'on the fly' with HACC simulations. A very fast halo finder [13] determines halo properties (both friends-

of-friends (FOF) and overdensity algorithms are supported) including halo mass and center, and halo profile and concentration. As described in Ref. [9] we now also have a GPU implementation of our halo finder. In addition to the fast halo finder, we have implemented a merger tree code that tracks the evolution of halos from early times to today and builds an evolution history for each halo. In Section 5.2 we present a fully portable version of our halo finder.

4 Code Development

Before the start of this project, the HACC framework only considered gravitational forces. The code has been optimized to achieve high-performance on a range of architectures, including BG/Q systems [5] and GPU enhanced systems [6]. For the KNL implementation, we started with the gravity-only version of HACC that has been previously optimized for the BG/Q systems. This version of the code uses a tree algorithm for the short-range solver and an FFT-based solver for the long-range forces. The new development that was carried out as part of this ESP project is the implementation of a hydrodynamics solver and a first set of sub-grid models into HACC, creating a new code, CRK-HACC. In the following, we report on the new developments leading to CRK-HACC and the performance and porting of HACC itself to the KNL architecture. We also discuss some of our efforts to develop fully portable analysis tools by employing the Thrust library.

4.1 New Methods/Algorithms: CRK-HACC

The CRK-SPH formalism has been incorporated into the gravity-only version of HACC to simulate the evolution of a baryonic fluid subject to both gravitational and hydrodynamic forces. Hydrodynamic forces are evaluated by generating a set of interaction lists for each CRK-SPH particle and sending these to architecture-optimized kernels. The interaction lists are built by walking an RCB (recursive coordinate bisection) tree whose structure is based off a variation of the tree currently used in the gravity solver. OpenMP threads are used to dynamically perform this task in parallel off a work queue of leaf nodes. Since the time steps demanded by the hydro solver are generally more restrictive than those required by gravity, we have devised a method to hierarchically time step the hydrodynamic force evaluation within each gravity step. We refer to the full implementation of CRK-SPH into HACC as CRK-HACC. While the code is designed to operate with both the gravity and hydro solvers activated, it maintains the flexibility for the user to specify running in either pure gravity-only or hydro-only mode. The latter has been used to validate the CRK-SPH framework within HACC, as described below.

4.1.1 Multi-species Initial Conditions in HACC

The first step in achieving hydrodynamic capabilities in HACC is to extend the use of the gravity-only code to properly coevolve multiple particle species. In cosmological applications, the two species of interest are cold dark matter (CDM) and baryons. From a gravitational standpoint, the two species in principle differ by (i) their mass and (ii) their initial power spectra. Obtaining the correct evolution of each component thus requires an understanding of any artificial coupling that may occur in the gravity solver between particles of unlike mass and/or initial power spectra. For this purpose, we have run a series of tests to systematically measure the amount by which each of these factors can impact the growth of structure formation in a multi-species simulation.

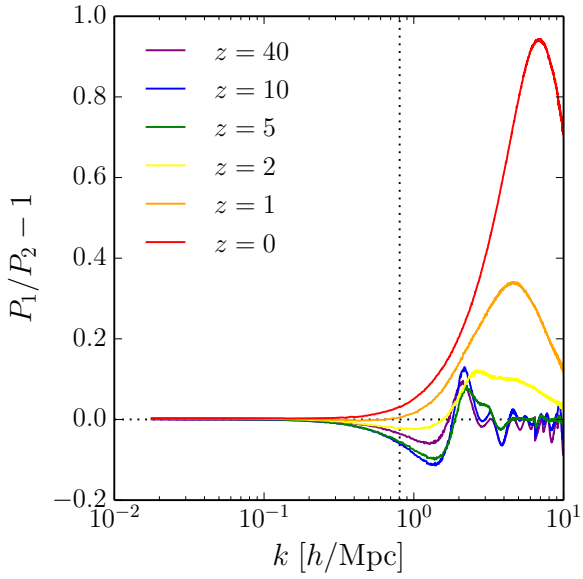

Figure 3: Power spectra of an unequal-mass multi-species simulation.
We show the fractional difference between the power spectra P_1 and P_2 of species 1 and 2, respectively, that have been initialized such that the mass $m_1 = 4m_2$. The initial power spectrum of each species is the same. The colored curves denote a given redshift between $z = 40$ and $z = 0$. The vertical dotted line denotes half the particle Nyquist frequency above which measurements of the power spectrum should not be analyzed.

We begin our tests by considering a two-component simulation where each species is initialized using a common power spectra, but are given different masses. In particular, species 1 is made four times as massive as species 2; a mass ratio similar to that of CDM and baryons. The particles are initialized in a box of width 500 h^{-1}Mpc at redshift $z = 200$ on a staggered lattice so that there are 256^3 particles in each species. We allow the simulation to evolve and measure the power spectra of each component separately at a series of redshifts between $z = 40$ and $z = 0$. In Figure 3 we show the evolution in the fractional difference of the power spectra of the two species. Note that if the two species were given the same mass, symmetry in the initial condition would require that $P_1 = P_2$ at all times. Hence, the deviations seen in Figure 3 are a numerical artifact sourced only by having unequal masses. If we analyze the power spectra individually, we find that at early times the massive (light) species develop a small suppression (enhancement) in power at small scales compared to the equal-mass case. This causes the suppression seen in Figure 3 for $z > 2$. The interpretation is that, compared to the equal-mass case, the gravitational potential around massive particles is somewhat enhanced, which causes nearby light particles to clump more within the enhanced potential. Reversing the argument can be used to explain the small-scale decrease in power for massive particles. At later times, however, the smallest scales enter the non-linear regime of structure formation and this seems to flip the relation around, leading to an enhancement in Figure 3 for $z < 2$. Nevertheless, the results shown here indicate that the difference in power between the two species is relatively small at all times. In fact, the power spectra agree at all times to within $\sim 5\%$ for $k < k_{\rm Nyq}/2$. Here $k_{\rm Nyq}$ is the Nyquist frequency of the particles and the limit $k_{\rm Nyq}/2$ denotes the maximum k effectively resolved by the simulation.

Next we consider a case where particles are initialized with the same mass but different power spectra. This was achieved by choosing a cosmology for which $\Omega_c = \Omega_b$ (i.e., CDM and baryons

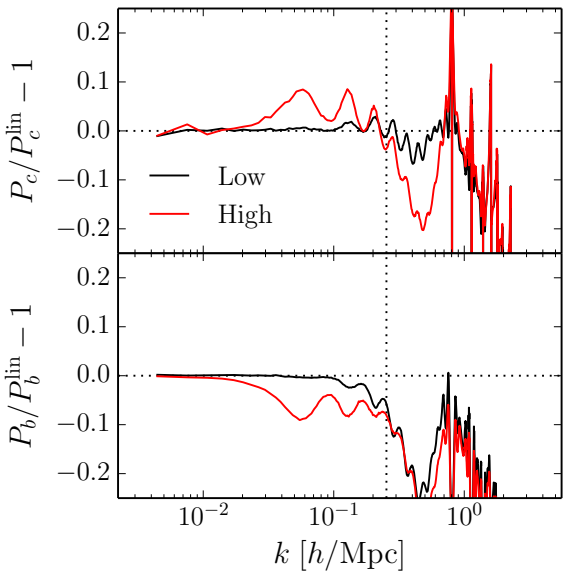

Figure 4: Power transfer within an equal-mass multi-species simulation. Fractional difference between the power spectrum of each species and its linear theory expectation at $z = 40$. The top panel corresponds to CDM while the lower panel shows baryons. The vertical dotted line in each panel denotes half the particle Nyquist frequency. The solid red lines correspond to our fiducial high-resolution simulation with a force resolution of 0.01Δ while the solid black lines denote a low-resolution simulation with a force resolution of $\sim \Delta$. Here Δ is the mean interparticle separation. The oscillations seen in each case are an imprint of the baryon acoustic oscillations present within the initial baryon transfer function.

contribute equally to the cosmological matter density) meaning that initializing with an equal number of CDM and baryon particles yields a common mass. The corresponding linear theory transfer functions for CDM and baryons were computed at the initial redshift of $z = 200$. We then allowed the simulation to evolve ten times steps, to $z = 40$, and compared the power spectrum of each component to its linear theory prediction. We focus on such an early redshift to ensure that all k modes are well within the bounds of linear theory. The solid red lines in Figure 4 show the fractional difference between the power of each component and its linear theory prediction at $z = 40$. For convenience, the curves have been smoothed on large scales by dividing out the initial fluctuations of the white noise field at $z = 200$. We see a systematic enhancement (suppression) in power within the range $10^{-2} \leq k \leq k_{\rm Nyq}/2$ for CDM (baryons). This has already developed to the level of $\sim 5\%$ after ten time steps and has propagated to relatively large scales. Hence, there appears to be some transfer of power between the two species based on having different initial power spectra. We have found that this effect can be diminished by decreasing the force resolution of the simulation. For example, the solid black line in Figure 4 shows the corresponding result when a low-resolution simulation is run by shutting off the short-range gravitational force calculation. In this case, the effective force resolution is roughly equal to the mean interparticle separation which significantly reduces coupling between nearby particles and alleviates power transfer between species. We note that the case considered here where CDM and baryons are initialized with their respective transfer functions is generally only important in cosmological applications where high-redshift results are desired. When focusing on low-redshift results, the power transfer shown here can be avoided by initializing each species to the total matter transfer function.

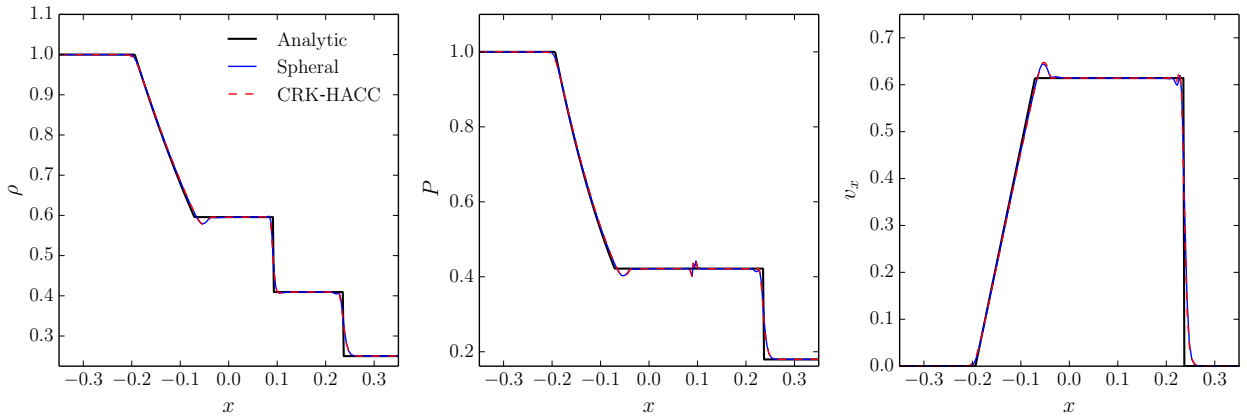

Figure 5: CRK-HACC 1D sod test.
Density (left), pressure (center), and velocity (right) as a function of position at time $t = 0.15$ for the 1D Sod shock tube test. The analytic calculation is traced by the solid black line while the solid blue and red dashed lines denote the CRK-SPH simulation results obtained from Spheral and CRK-HACC.

These tests demonstrate that one must be careful to consider artificial coupling between unlike species. This can occur from having either unequal particle masses or differing initial transfer functions. In either case, the coupling can be reduced by smoothing the force resolution around unlike particles. One possibility is to implement an adaptive smoothing length for unlike particles that is based on the local density. This is a subject of ongoing work.

4.1.2 CRK-SPH Tests on KNL

We have extensively tested the non-cosmological CRK-SPH implementation within HACC on the KNL architecture. We have validated the results of the hydro solver by running hydro-only tests and comparing the output to analytic predictions and to CRK-SPH implemented within the LLNL-developed Spheral framework.

The first case we consider is the Sod shock tube test. This is a 1D problem for which a box of unit length is initialized with two uniform gaseous components. More specifically, a high-density, high-pressure gas is placed on the left side of the box ($-0.5 \leq x \leq 0$) while a low-density, low-pressure gas occupies the right ($0 \leq x \leq 0.5$). The density and pressure of the two mixtures are $(\rho_{\rm high}, P_{\rm high}) = (1, 1)$ and $(\rho_{\rm low}, P_{\rm low}) = (0.25, 0.1795)$ with each component having zero initial velocity. In this case, we achieve a uniform particle mass by starting with four times the number of particles in the high-density region than the low-density region; we use here particle counts $(N_{\rm high}, N_{\rm low}) = (400, 100)$. The two gases begin interacting at time $t = 0$ leading to the propagation of a shock front into the low-density region and a rarefaction wave in the high-density region. It is easy to analytically calculate the state of the gas as a function of position within the box at all later times.

In Figure 5 we show the density, pressure, and velocity of the gas at time $t = 0.15$. At this point the leading edge of the rarefaction wave has propagated to $x \approx -0.2$ while the shock front has advanced to $x \approx 0.25$. The plots are centered at the location of the original interface at $x = 0$. The solid black line in each panel shows the analytic calculation while the solid blue and dashed red lines trace the results from Spheral and CRK-HACC, respectively. Both simulations agree well with the

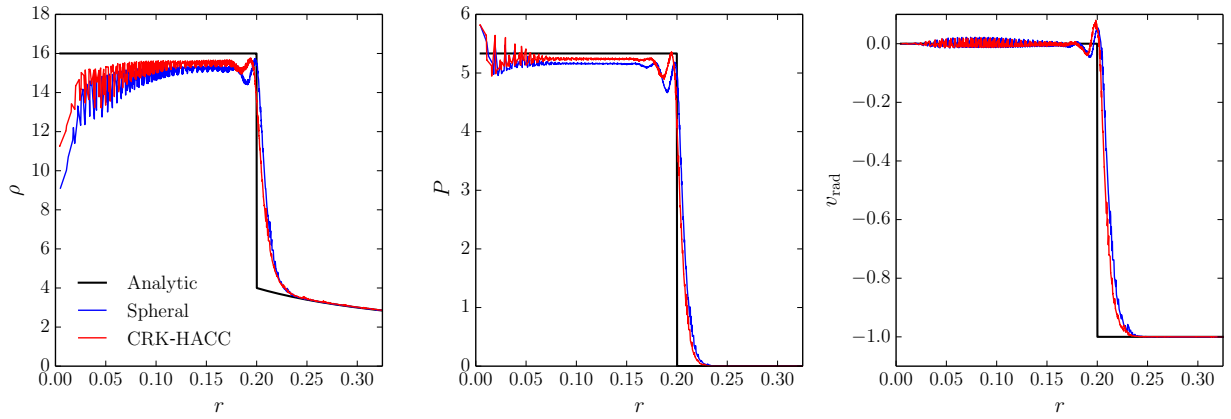

Figure 6: CRK-HACC 2D noh test.
Density (left), pressure (center), and radial velocity (right) as a function of radial distance from the origin at time $t = 0.6$ for the 2D Noh implosion test. The analytic calculation is traced by the solid black line while the solid blue and red dashed lines denote the CRK-SPH simulation results obtained from Spheral and CRK-HACC.

analytic result and we see an excellent match between Spheral and CRK-HACC. Most notable is the ability of each code to smoothly capture the sharp shock front. We have also performed the Sod test in two and three dimensions by stacking the initial condition in y and z and find the same general agreement shown here.

The next case we consider is the Noh implosion test. In this test, we start with a uniform gas of unit density and zero pressure inside a box of unit length. Each gas particle is initialized with a velocity directed radially toward the center of the box with a magnitude of one. This setup leads to the development of a shock front that propagates radially outward from the origin. The Noh test can be performed in any number of dimensions and has an analytic solution for each case. We consider here a 2D implementation where 100×100 particles are initially arranged on a uniform lattice at time $t = 0$.

Figure 6 shows the density, pressure, and radial velocity of the gas at time $t = 0.6$. At this point, the shock front has advanced to a radial distance of $r_s \approx 0.2$. The analytic solution predicts gas in the pre-shock region ($r < r_s$) to have $\rho = 16$, $P = 16/3$, and $v_{\rm rad} = 0$. This is shown by the solid black line in Figure 6 while the solid blue and red lines trace the radial profiles obtained from Spheral and CRK-HACC, respectively. We see that both codes do extremely well at resolving the sharp shock front and yielding the pre-shock analytic values. We do not expect the two codes to agree exactly since there are minor differences in their methods (e.g., time-stepping schemes, Spheral is run in double precision while HACC is in single precision) and, in this case, we see that CRK-HACC does slightly better at matching the analytic result. Note that the zigzags seen in the radial profiles are simply an imprint of the initial lattice configuration and can be removed by alternatively placing particles in a radially symmetric configuration or using a glass initial condition. The drop-off observed in density at small radii is caused by difficulty in resolving the initial singularity in the radial velocity field at the origin. This numerical artifact, known as wall-heating, is substantially alleviated in the CRK formalism compared to traditional SPH methods. We have also performed the Noh test in 3D and find consistent results with the 2D case present here.

The final test we consider is the 3D Blob test. In this problem, we start with a spherical cloud, or

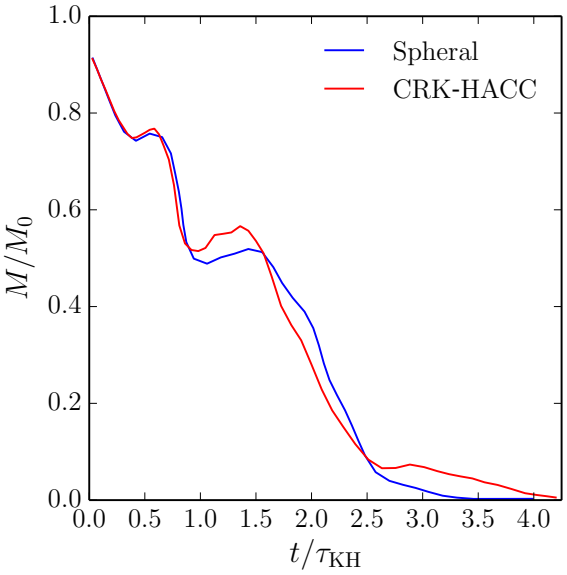

Figure 7: CRK-HACC 3D blob test.
Mass of the blob as a function of time for the 3D Blob test. Here mass is normalized to its initial value while time is normalized to the Kelvin-Helmholtz mixing time. The solid blue and red lines trace the results obtained from Spheral and CRK-HACC, respectively.

"blob", of density $\rho_c = 10$ situated within a uniform background medium of density $\rho_b = 1$. The center of the cloud is placed at $(x_c, y_c, z_c) = (5, 5, 5)$ with a radius $r_c = 1$ while the background material fills the volume contained in the rectangular box of length $(L_x, L_y, L_z) = (10, 10, 40)$. Both the cloud and background material are initialized with pressure $P = 1$. The cloud is initially at rest while the background material is given a constant velocity $v_z = 2.7c_s$ where c_s is the sound speed of the gas. This leads to the development of a bow shock on the leading face of the blob followed by a gradual evaporation of its contents as the supersonic background material shears past and turbulently mixes with the cloud. We consider here a case for which particles are initialized on a uniform mesh with separation $dl = 10/128$ in each dimension to fill the background medium. Particles falling within the initial volume of the blob are excised from this lattice and replaced by a particle packing that is ρ_c/ρ_b times finer than the background.

Figure 7 shows the mass contained within the blob as a function of time. The mass is normalized to its initial value, M_0, while time is normalized to the Kelvin-Helmholtz mixing time-scale, $\tau_{\rm KH}$. The solid blue and red lines trace the results obtained by Spheral and CRK-HACC, respectively. In this case, there is no analytic result to compare against. As before, we find good agreement between the two codes. This is especially encouraging given that the blob test is relatively sensitive to initial conditions, which were not completely equivalent between Spheral and CRK-HACC. Since HACC requires a cubic periodic domain, the blob test was tiled four times in x and y to fill the entire volume, thus its evolution is not completely identical to the rectangular reflective domain used in Spheral. Nevertheless, both codes yield quantitatively similar results and both achieve full evaporation of the cloud. This highlights the full utility of the CRK-SPH formalism as traditional SPH methods fail to fully evaporate the blob due to their inability to properly mix the two fluids.

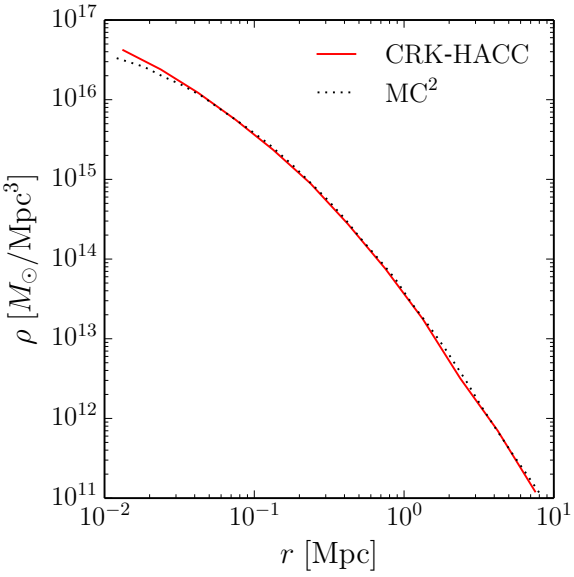

Figure 8: Radial density profile for the gravity-only Santa Barbara test.
The solid red line traces the radial density profile for the total matter (CDM plus baryon) component of the main Santa Barbara cluster at $z = 0$. The dotted black line shows the single-species gravity-only run of Ref. [10] which was performed using MC^2 (the predecessor code to HACC). Our agreement with the dotted black line verifies that the gravity solver in CRK-HACC is properly evolving multiple species.

4.1.3 Santa Barbara Test

The correct implementation of the full gravity plus CRK-SPH hydrodynamics capability has been tested by running the Santa Barbara cluster comparison benchmark for hydrodynamic cosmology codes. This test was first introduced in Ref. [2] where 12 different codes were run on the same initial conditions to simulate one massive cluster ($\sim 10^{15}$M$_\odot$) in a small cosmological volume including an adiabatic hydrodynamic treatment. Several cluster properties were then measured and compared, including global properties such as mass as well as radial profiles of the gas and dark matter density. While the test is moderate in size for today's standards with 2×256^3 particles, it is still a very useful test since a lot of comparison data from different algorithmic implementations from AMR to SPH is available.

We begin with an investigation of a gravity-only version of the Santa Barbara test. This is useful for comparison purposes against the full test to highlight the effects of hydrodynamics. Figure 8 shows the radial density profile of the total cold dark matter (CDM) plus baryonic component of the main cluster at $z = 0$. This is compared against the radial profile of Ref. [10] which performed a similar gravity-only run except containing only the CDM particles. Hence, the fact that we are able to reproduce the total radial density profile here while simulating both CDM and baryons confirms that the gravity solver is properly working within the multi-species framework of CRK-HACC.

The full hydrodynamic Santa Barbara simulation is currently running and has evolved to $z = 1$ at the time of writing. We present here some preliminary results from the run. Firstly, in Figure 9 we compare the power spectra of each species in the hydro simulation to the corresponding result in the gravity-only run. We find the expected result that the baryonic power spectrum is heavily

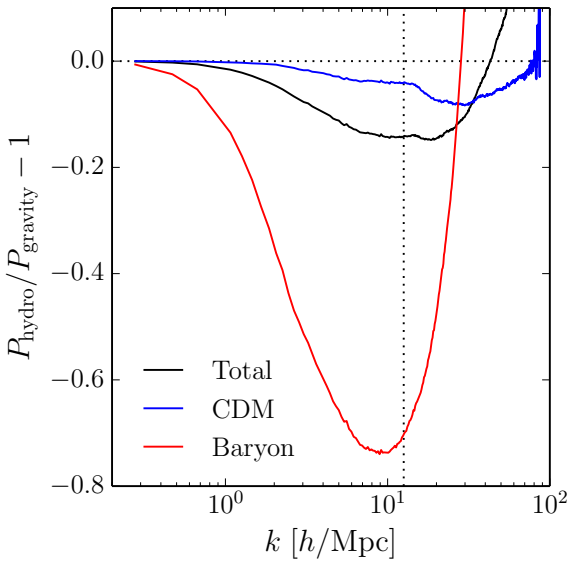

Figure 9: Santa Barbara Power Spectra
The fractional difference in power at $z = 1$ for CDM (blue), baryons (red), and the total matter (black) between the full hydrodynamic Santa Barbara simulation and its gravity-only analog. The vertical dotted line denotes half the particle Nyquist frequency specifying the maximum k resolved in the simulation.

suppressed due to the presence of thermal pressure which resists gravitational collapse. This leads to a small backreaction on the CDM which also displays moderate suppression in power on the smallest scales. Of course, the total matter power sits between the two regimes with $\sim 15\%$ suppression at the largest k. In Figure 10 we visualize the suppression in baryon clumping by comparing slices of the CDM and baryon density fields. Each slice is centered on the location of the most massive halo in the box at $z = 1$ and shows density normalized to the mean density of each species. It is easy to see by eye the smoothing of the inner portion of the main halo due to baryonic thermal pressure. In general, the baryon density slices are smoother and more diffuse around the location of each dense CDM halo.

4.1.4 Subgrid Model Development

We have implemented a first set of sub-grid models into CRK-HACC on the KNL architecture. In our model selection we follow loosely Ref. [1], a paper that describes simulations of the Sunyaev-Zel'dovich power spectrum with AGN feedback. The sub-grid models include a UV background following Ref. [3], a stellar formation model both for star formation and corresponding galactic winds [12], and the AGN model first described in Ref. [1].

Implementation of the Homogeneous UV Background Model

We have implemented a homogeneous radiative cooling model following Ref. [11], where we assume a spatially uniform, but time-varying ultraviolet background (UVB) radiation field from Ref. [4]. All particle (cell) volumes are assumed to be optically thin, and radiative feedback is accounted for via the UVB model; we currently do not follow radiation transport through the box, nor do we explicitly account for the effects of thermal feedback of stars, quasars, or active galactic nuclei. In

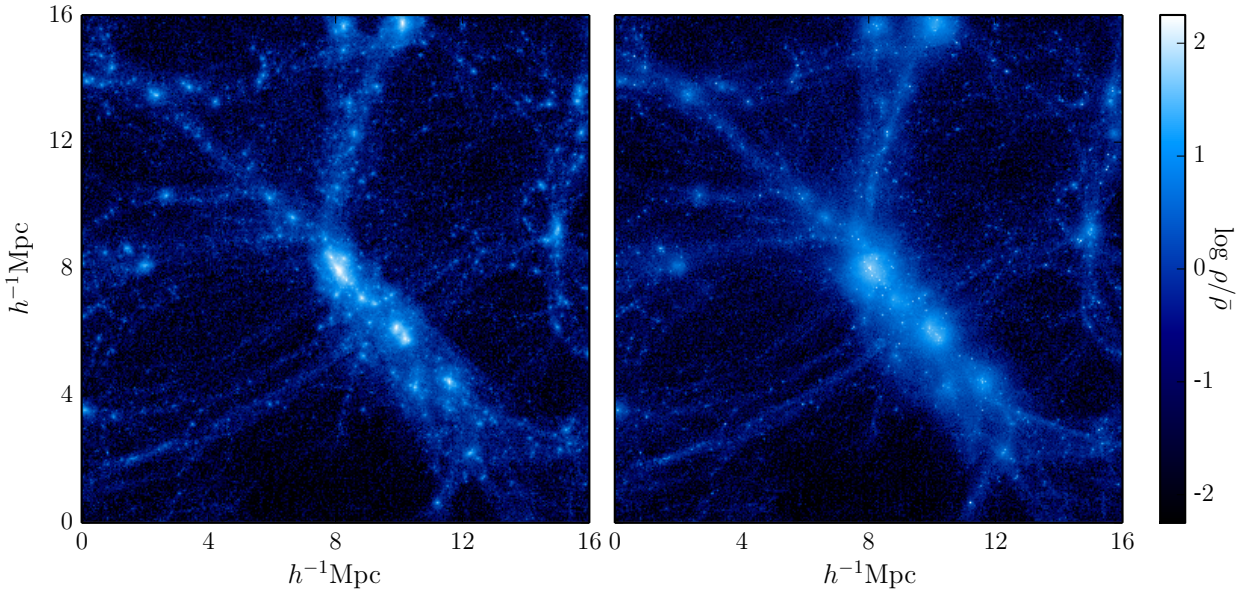

Figure 10: Santa Barbara Density Slices
We show slices of the CDM (left) and baryon (right) density fields centered on the location of the most massive halo in the hydro run at $z = 1$. Each slice has a width of 16 h^{-1}Mpc (half the extent of the full box) and projects through a distance of 6.4 h^{-1}Mpc.

addition, we include inverse Compton cooling off the microwave background.

In more detail, we model the chemistry of the gas as having a primordial composition with hydrogen and helium. The reaction network includes 6 atomic species: H I, H II, He I, He II, He III and e^-, which we evolve under the assumption of ionization equilibrium. The resulting system of algebraic equations is:

$$\left(\Gamma_{e,\mathrm{H\,I}} n_e + \Gamma_{\gamma,\mathrm{H\,I}}\right) n_{\mathrm{H\,I}} = \alpha_{\mathrm{r,H\,II}} n_e n_{\mathrm{H\,II}} \tag{1}$$

$$\left(\Gamma_{e,\mathrm{He\,I}} n_e + \Gamma_{\gamma,\mathrm{He\,I}}\right) n_{\mathrm{He\,I}} = \left(\alpha_{\mathrm{r,He\,II}} + \alpha_{\mathrm{d,He\,II}}\right) n_e n_{\mathrm{He\,II}} \tag{2}$$

$$\left[\Gamma_{\gamma,\mathrm{He\,II}} + \left(\Gamma_{e,\mathrm{He\,II}} + \alpha_{\mathrm{r,He\,II}} + \alpha_{\mathrm{d,He\,II}}\right) n_e\right] n_{\mathrm{He\,II}} \tag{3}$$

$$= \alpha_{\mathrm{r,He\,III}} n_e n_{\mathrm{He\,III}} + \left(\Gamma_{e,\mathrm{He\,I}} n_e + \Gamma_{\gamma,\mathrm{He\,I}}\right) n_{\mathrm{He\,I}} \tag{4}$$

combined with three closure equations for the conservation of charge and hydrogen and helium abundances. Radiative recombination ($\alpha_{\mathrm{r,X}}$), dielectronic recombination ($\alpha_{\mathrm{d,X}}$), and collisional ionization ($\Gamma_{e,\mathrm{X}}$) rates are strongly dependent on the temperature, which itself depends on the ionization state through the mean mass per particle μ

$$T = (\gamma - 1)\frac{m_p}{k_{\mathrm{B}}}\mu\ e_{\mathrm{int}} \tag{5}$$

where m_p is the mass of a proton, k_{B} is the Boltzmann constant, and e_{int} is the internal thermal energy per mass of the gas. For a gas composed of only hydrogen and helium, μ is related to the number density of free electrons relative to hydrogen by $\mu = 1/\left[1 - (3/4)Y + (1 - Y)n_e/n_{\mathrm{H}}\right]$, where

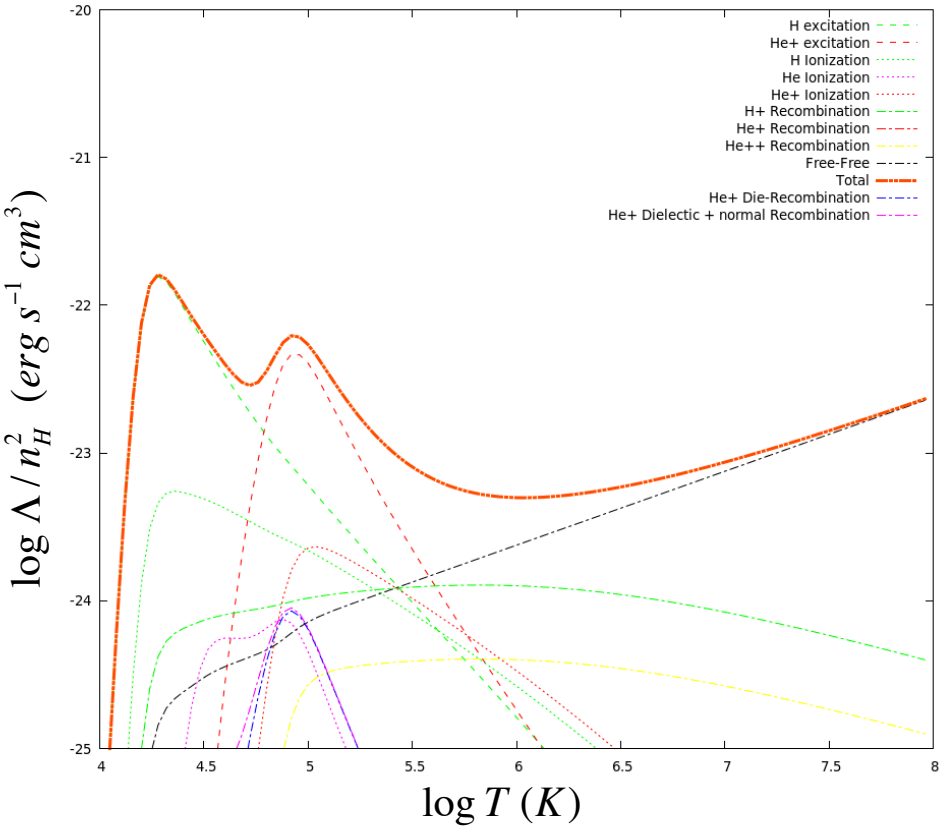

Figure 11: UV Cooling Rates as Function of Temperature
Calculated UV cooling rates (normalized) vs. temperature for the test case of zero photoionization.

Y is helium mass abundance. We iteratively solve the reaction network equations together with the ideal gas equation of state, $p = (\gamma - 1)\rho e_{\rm int}$, to determine the temperature and equilibrium distribution of species. Figure 11 plots the resulting normalized cooling rates as a function of temperature in the special case of no photoionization, which matches the results found in Ref. [11] (Figure 1 in that publication).

Implementation of AGN Feedback

Active galactic nuclei, or AGN, are the core of very luminous galaxies with a broad electromagnetic spectrum, called active galaxies. For example, quasars, which are a class of AGN, are among the brightest objects of the Universe and they can emit radiation at all frequencies from radio to X-rays and even gamma rays. This indicates that unlike other galaxies, radiation from an active galaxy is emitted by a non-thermal process and is thus not the sum of the radiation due to the stars in the galaxy. There is growing evidence that the center of these active galaxies is populated by a super-massive black hole (from 10^6 to $10^9 M_\odot$). AGN are classified under several classes according to their observational properties but it is believed that for any of them radiation is a consequence of an accretion process as summarized below:

1. Gas forms an accretion disk and falls into a central super-massive black hole.
2. A fraction of the accreted gas is ejected under the form of relativistic jets of particles (feed-

back).

3. The surrounding gas interacts with the jets and energetic radiation is emitted.

Because of the gravitational influence and the processes of accretion and feedback, black holes at the centers of active galaxies are believed to have a significant influence on the shape of the galaxies. Unfortunately, modeling AGN in large-scale cosmology codes such as HACC (and even in galaxy-zoomed simulations) is not straightforward. There two main reasons for that, the first one being the limited knowledge of the AGN feedback process. The second one is the fact that the AGN processes happen at scales lower than HACC's resolution. To still take the effects of AGN into account without correctly modeling AGN structures and processes, we use a subgrid model. The goal of a subgrid model is to reproduce the effects at large scales of an unresolved physical process.

HACC deals with two kinds of particles: dark matter and baryonic particles. A flag is set to distinguish baryonic particles from dark matter particles. We add a third type of particle, the black hole particle, which is a special baryon that does not interact with other baryons except to absorb them and give them energy. A halo finder is frequently launched and the most bound particle in the halo with a mass that exceeds the threshold $m_{\text{halo,min}}$ is marked with a special flag to indicate that there is now a black hole at the halo center. When the black hole is placed at the center of the halo, it is given a mass $M_{BH,0} = M_{seed}$. The internal mass of the black hole is stored in a separate array from the actual mass of the baryonic particle used for the gravity computation. All AGN computations are done at short range, and use the tree built for gravitational or hydrodynamical forces. We have implemented four kernels to capture AGN physics which we briefly describe in the following.

`AGN_UPDATE` The purpose of this kernel is to compute at each step the internal mass M_{BH} of each black hole. We rely on the Bondi-Hoyle formula

$$\dot{M}_{Bondi} = \frac{4\pi\alpha G^2 M_{BH}^2 \rho}{(c_s^2 + v^2)^{\frac{3}{2}}}. \tag{6}$$

The local density of gas ρ is computed via an SPH-kernel, as well as the local internal energy u, which is used to compute the speed of sound $c_s = \sqrt{\gamma(\gamma-1)u}$, where γ is the heat capacity ratio of the gas (considered ideal).

We then update the internal mass and the amount of energy that needs to be spread to the nearby gas. Note that we subtract the fraction of the accreted matter converted into feedback energy.

$$M_{BH} = M_{BH} + (1-\epsilon_r)\dot{M}\Delta t \tag{7}$$

$$E_{feed} = \dot{E}_{feed} + \varepsilon_f \varepsilon_r \dot{M}_{BH} c^2 \tag{8}$$

`AGN_ACCRETION` This kernel browses all the baryonic particles of the interaction list of each black hole particle. For each of them, it uses the specific criteria to decide if the particle should be absorbed by the black hole or not. If it is the case, the momentum of the particle is added to the black hole and the particle is deleted. Deleting particles is somewhat tricky because the tree data structure relies on the order of the particles in their containing array. Thus, particles are not deleted immediately but their identifier is stored for later deletion. For now, only stochastic-unconditional accretion is implemented.

`AGN_MERGE` This kernel browses all of the black hole particles in the interaction list of each black hole and decides if the two black holes should merge, according the criteria in **??**. Thus, it is very similar to the `AGN_ACCRETION` kernel. However, since leaves are browsed in parallel, we have to make sure that the same merging decision and order is taken for all black holes (especially if a black hole can merge with more than one particle). All the merging decisions are logged in a shared data-structure, so that each thread can check it before making any merging decision.

`AGN_FEEDBACK` This kernel computes the amount of energy received by the particles in the neighborhood of black hole particles. For now, the approach described in Ref. [12] is implemented.

All the kernels have been implemented in HACC on the KNL architecture and have been thoroughly tested. With these new sub-grid model we can now carry out new simulations with CRK-HACC exploring how baryonic physics and feedback influence structure formation in the Universe.

4.2 Optimization for Theta

After having provided an extensive discussion about the new models and physics that have been implemented into HACC to create the new CRK-HACC we now turn to the performance HACC and CKR-HACC achieve on Theta. We start by discussing on-node performance results and then show strong and weak scaling performance.

4.2.1 On-node Performance of HACC on the KNL

We present comparisons of KNL and BG/Q performance in Table 1 and Table 2. Table 1 employs one node of the BG/Q (16 cores) versus a quadrant of the KNL (16 cores), while Table 2 shows results from four BG/Q nodes versus one KNL node. The two test cases vary slightly in size. In each case, the number of ranks per node and threads per rank are varied. The individual-core speed-up of the KNL over the BG/Q ranges from close to a factor of five to as high as a factor of eight. The fastest time-to-solution for both architectures is obtained with 8rpn (ranks per node) and 8 OpenMP threads/32rpn, in which case the ratio of the KNL to BG/Q performance is roughly five.

An important point to note about the HACC performance numbers is the high level of computational intensity in the short-range gravitational force kernel. For this reason, operating the KNL in its three different memory access modes: DDR only (flat mode), MCDRAM (or HBM) only, and cache mode (MCDRAM used as a DDR cache) makes no difference; the timings are essentially identical in all cases.

Next we show a comparison of HACC performance on KNL versus Haswell X86 nodes. Results are presented in the same manner as the first comparison with the BG/Q, shown in Table 1. (This test compares 16 KNL cores vs. 16 Haswell cores.) At a core vs. core level, running on Haswell shows a clear performance advantage even though no attempt was made to fully optimize HACC for this architecture (except running with AVX2). We note, however, that a core vs. core comparison of KNL vs. Haswell is not a fair metric for a compute-dominated code like HACC since individual Haswell cores are significantly faster than their KNL counterparts.

Cores	RPN	OMP	TH	BG/Q [Time/s]	KNL B0 [Time/s]	Ratio
16	4	4	16	4297	616.3308	6.98
16	4	8	32	2677	543.7294	4.92
16	4	16	64	2504	530.2267	4.72
16	8	2	16	4362	544.7519	8.00
16	8	4	32	2571	459.5265	5.59
16	8	8	64	**2278**	**437.2058**	5.18
16	16	4	64	2581	468.5037	5.50

Table 1: HACC timing comparing one BG/Q node and one KNL quadrant.
One BG/Q node (16 cores) compared to one KNL quadrant (16 cores); RPN in the Table denotes the number of MPI ranks for each run, it is not literally the number of ranks per node. The corresponding number of OpenMP threads per MPI rank is denoted by OMP, and the total number of threads by TH. Problem setup: $np = ng = 320$ (where np^3 is the number of particles and ng^3, the number of grid points), 3 full time steps, 5 sub-steps per time step. Results from the fastest configuration are highlighted in bold face.

Cores	RPN	OMP	TH	BG/Q [Time/s]	KNL B0 [Time/s]	Ratio
64	16	4	64	4542	678.7571	6.69
64	16	8	128	2823	606.1815	4.66
64	16	16	256	2556	587.2716	4.35
64	32	2	64	4747	620.7261	7.65
64	32	4	128	2824	536.1650	5.27
64	32	8	256	**2503**	**503.0927**	4.98
64	64	4	256	2539	510.3745	4.98

Table 2: HACC timing comparing four BG/Q nodes versus one KNL node.
One KNL node (64 cores) versus four BG/Q nodes (64 cores). Problem setup: $np = ng = 512$, 3 full time steps, 5 sub-steps. Conventions are the same as in Table 1

4.2.2 On-node Performance of CRK-HACC on the KNL

In this section we present comparisons of CRK-HACC to the HACC gravity-only solver on KNL. Tables 4 and 5 show the timing of CRK-HACC for the single quadrant and single node KNL problem setups shown previously for the HACC gravity-only solver. In each setup, the total number of particles is doubled in CRK-HACC compared to the gravity-only case to accommodate the additional set of CRK-SPH particles that experience both gravitational and hydrodynamical force evaluations. Both setups show that the runtime of the CRK-HACC simulations are roughly a factor of 40 times that of their gravity-only counterparts. Half of this is due to the factor of two increase in the particle count while the remaining is sourced in the hydro solver which contains considerably more computations than the gravity solver. Interestingly, we find that the fastest time-to-solution in CRK-HACC is achieved when only two hardware threads per core are activated compared to the case of four threads per core seen in the gravity-only tests. Additional tests show that, as was the case with the gravity solver, the CRK-HACC hydro kernels are insensitive to the choice of memory configuration on the KNL; similar speeds are achieved in cache mode to both DDR-only and MCDRAM-only flat modes. Further work on hydro kernel performance optimization is continuing.

KNL Bin3 64 cores				Haswell, 2x18 cores			
RPN	OMP	TH	Time/s	RPN	OMP	TH	Time/s
4	4	16	616.3308	4	4	16	391.7491
4	8	32	543.7294	4	8	32	349.8340
4	16	64	530.2267				
8	2	16	544.7519	8	2	16	357.9577
8	4	32	459.5265	8	4	32	**311.8854**
8	8	64	**437.2058**				
16	4	64	468.5037	16	2	32	335.8463

Table 3: HACC timing comparing KNL Bin3 and Haswell Xeon E5-2699 v3
Problem set up: $np = ng = 320$, 3 full steps, 5 substeps. We mark the fastest set-up in bold face.

Cores	RPN	OMP	TH	Gravity-only [Time/s]	CRK-HACC [Time/s]	Ratio
16	4	4	16	616.3308	21881	35.5
16	4	8	32	543.7294	18598	34.2
16	4	16	64	530.2267	19685	37.1
16	8	2	16	544.7519	21750	39.9
16	8	4	32	459.5265	**18548**	40.3
16	8	8	64	**437.2058**	20026	45.8
16	16	4	64	468.5037	21339	45.4

Table 4: CRK-HACC timing compared to gravity-only on one KNL quadrant.
The problem setup of Table 1 is repeated for CRK-HACC on one KNL quadrant (16 cores). The numbers shown in the fifth column are the timing of the HACC gravity-only solver shown previously. In the case of CRK-HACC the number of particles is doubled ($np = 2 \times 320$) while all other parameters are unchanged (i.e., $ng = 320$, 3 full time steps, 5 sub-steps per time step.) Results from the fastest configuration are highlighted in bold face.

4.2.3 Strong and Weak Scaling of HACC on the KNL

In this section we present scaling results obtained on Theta with the full HACC gravity-only code. We focus our tests on two simulation problems. The first is a strong scaling example, evolving $\sim 2048^3$ particles (the exact number depending on the problem decomposition) in an $(800h^{-1}\mathrm{Mpc})^3$ volume. This is the simulation size chosen for our final hydrodynamic simulation that is part of this project. The second concerns a high-resolution, large volume, gravity-only simulation relevant for, e.g., generating synthetic catalogs for the Large Synoptic Survey Telescope (LSST) data challenges. In order to obtain a good balance between mass resolution, volume, and wall-clock time for a full cosmological simulation, we set up a weak scaling test that would lead to a volume of $(2500h^{-1}\mathrm{Mpc})^3$, evolving 9216^3 particles on 3072 Theta nodes. This simulation has a particle mass of $m_p \sim 1.5 \cdot 10^9 h^{-1}\mathrm{M}_\odot$, depending on the exact cosmological parameters chosen.

The strong scaling test employs two simulations: 1) 2048^3 particles on 64 to 2048 nodes and 2) 2304^3 particles on 3072 nodes (this minor change is for decomposition reasons, due to the number of nodes on Theta). Each simulation was run with 8rpn (ranks per node). For the 3-D parallel FFT, we used an unthreaded version of FFTW for the node-local 1-D FFT. For the short-range solver, we used HACC's local-tree or "bush" version (compared to the earlier monolithic tree version in the HACC CORAL benchmark) and chose a chaining mesh size of 8.0, 16 threads per processor, -d=16, and -j=2. This code version gives us more flexibility to optimize the time to solution. Various

Cores	RPN	OMP	TH	Gravity-only [Time/s]	CRK-HACC [Time/s]	Ratio
64	16	4	64	678.7571	22785	33.6
64	16	8	128	606.1815	**21392**	35.3
64	16	16	256	587.2716	22303	38.0
64	32	2	64	620.7261	22915	36.9
64	32	4	128	536.1650	21500	40.1
64	32	8	256	**503.0927**	23628	46.9
64	64	4	256	510.3745	24997	48.9

Table 5: CRK-HACC timing compared to gravity-only on one KNL node
The problem setup of Table 2 is repeated for CRK-HACC on one KNL node (64 cores). The numbers shown in the fifth column are the timing of the HACC gravity-only solver shown previously. In the case of CRK-HACC the number of particles is doubled ($np = 2 \times 512$) while all other parameters are unchanged (i.e., $ng = 320$, 3 full time steps, 5 sub-steps per time step.) Results from the fastest configuration are highlighted in bold face.

parameters will need to be adjusted in the near future to yield the best possible time to solution. We use 96 particles per leaf node for all strong and weak scaling tests reported here. A first set of tests showed that this number is still close to optimal – as it was on the BG/Q – for the best time-to-solution. The results are shown in Fig. 12 in the left panel; we show timings for the 3-D FFT (which is a major component of the long-range force solver and dominates the scalability properties of HACC) and the total time per particle per step, which includes the short-range solver. HACC strong-scales remarkably well on Theta, promising good performance for our future hydrodynamics simulations as well.

Investigation of the weak scaling properties of HACC is driven by the requirement to carry out a large-volume $(2.5h^{-1}\text{Gpc})^3$ simulation with high mass resolution to be relevant for a wide and deep survey such as LSST. The tests carried out have a mass resolution of $m_p \sim 1.5 \cdot 10^9 h^{-1}\text{Mpc}$ – we adjust the volume of the simulation to the number of particles run to keep the mass resolution constant. In other words, we keep the volume and number of particles per rank approximately the same. This means that the smallest simulation on six nodes evolves 1152^3 particles in a $(312.5h^{-1}\text{Mpc})^3$ volume and the largest simulation evolves 9216^3 particles in a $(2500h^{-1}\text{Mpc})^3$. The final simulation was run on 3072 nodes. The general settings (number of ranks per node, threads, particle per leaf node, etc.) are the same as in the strong scaling tests. Again, for absolute time to solution these will be fine-tuned over time. The weak scaling results are shown in Fig. 12 in the right panel – following the previous trends for strong scaling, the weak scaling results are also excellent.

We have successfully run a larger FFT of size $11{,}520^3$ on Theta. Scaled up to the number of ranks that were available on Mira, the FFT is approximately a factor of two faster than on the BG/Q. Since we are interested in lowering the absolute time to solution for the FFT (which is a little less than a minute with the current settings on 3072 nodes), it is important to investigate threading options and different 1-D FFT implementations.

4.2.4 Strong Scaling of CRK-HACC on the KNL

For CRK-HACC we have set up the same strong scaling test as for HACC, described in Section 4.2.3. The only difference is that the CRK-HACC version evolves twice as many particles (2048^3 dark matter particles and 2048^3 baryons) in order to keep the particle mass consistent between the two

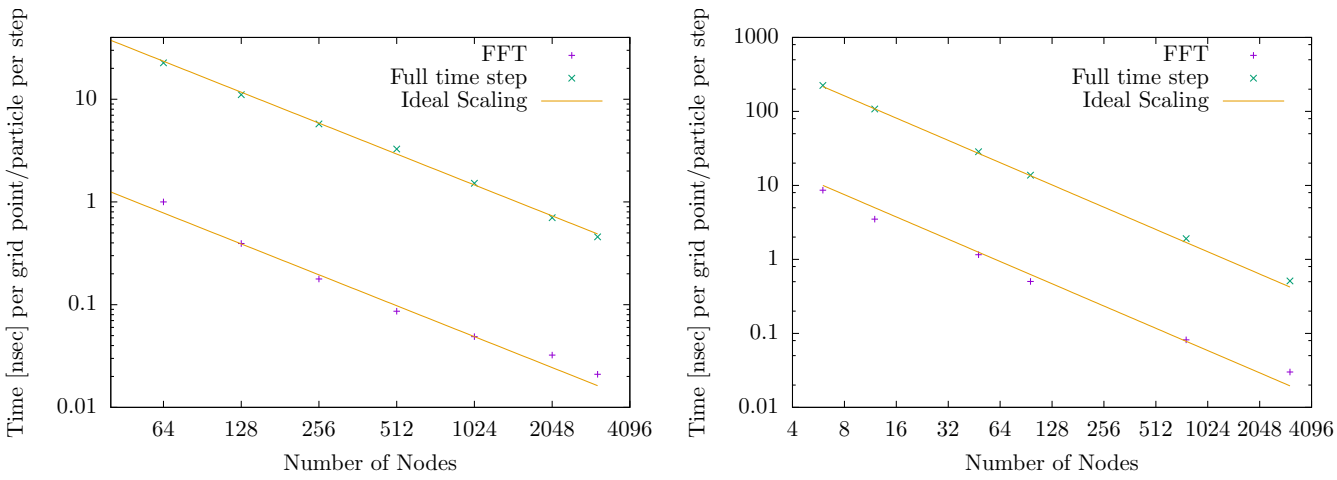

Figure 12: HACC Gravity-only strong and weak scaling results.
Left panel: Strong scaling results for the FFT (purple pluses) and the full high-resolution time stepper (green crosses); solid lines represent ideal scaling. We show the time per grid cell (FFT) or particle (full time step) per step. The tests were carried out to a maximum of 3072 nodes (almost 200k cores). The problem size was set to mimic the size of the hydrodynamics run proposed for our ESP simulation, though the timings as shown are only for the gravity solver. Right panel: Weak scaling results for the FFT (purple crosses) and the full high-resolution time stepper (green crosses). The solid line represents ideal scaling. As in Fig. 12 we show the time per grid cell (FFT) or particle (full time step) per step. The tests were carried out on up to 3072 nodes. The problem size was set to mimic a high mass resolution run ($m_p \sim 1.5 \cdot 10^9 h^{-1}$Mpc) in a large volume. We kept the mass resolution constant by tuning particle number and simulation volume. For this test, we kept the size of the problem per rank the same and scaled up the problem size with more ranks.

runs. In the case of HACC, the particle carries the combined mass of dark matter and baryons, while in CRK-HACC the two particle species in the code have different masses that when combined result in the HACC particle mass. The result for the strong scaling test is shown in Fig. 13 including the HACC gravity-only scaling results. The strong scaling behavior of CRK-HACC is essentially perfect. We show the time spent per particle (active and passive) for a full time step accounting for the fact that CRK-HACC evolves twice as many particles. Overall, CRK-HACC is approximately 16 times slower than HACC per particle, consistent with the significantly increased computational load due to the addition of CRK-SPH.

5 Portability

5.1 HACC and CRK-HACC

By its design, HACC (and therefore CRK-HACC) is targeted to run on all available HPC architectures and does so currently. The project philosophy for portability is as follows. Certain code changes to achieve good performance are inevitable, but must be limited in scope and should not hurt code flexibility, maintenance, and extensibility. HACC has this feature built into its design (95% of the code remains unchanged even when radically different architectures are used). The work needed to include a new supported platform is therefore not excessive. This was the case for

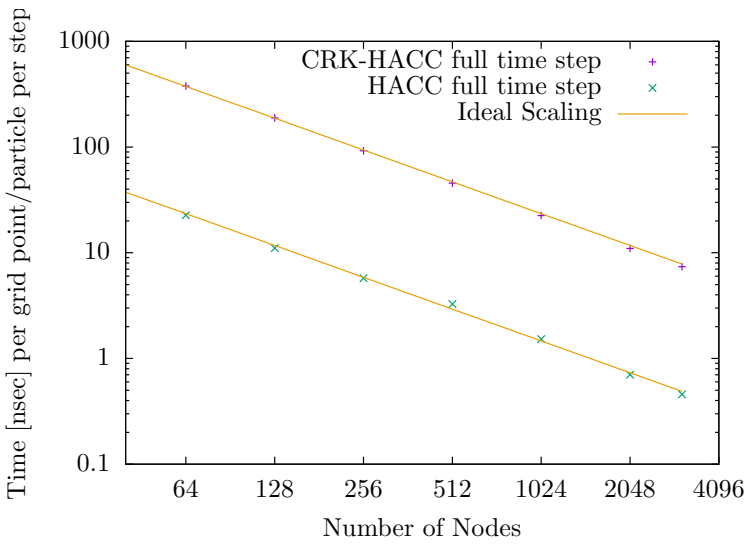

Figure 13: CRK-HACC strong scaling results.
Strong scaling results for CRK-HACC (purple crosses) and HACC in gravity-only mode (green crosses) for comparison. The solid lines represent ideal scaling. The tests were carried out on up to 3072 nodes. The set-up of the problem is exactly the same as shown in Fig. 12. The final point on 3072 nodes shows the set up for our first hydro run on Theta. The CRK-HACC version is roughly 16 times slower than HACC gravity-only, better than predicted by initial estimates.

our port to Theta. Our team carried out some minor optimization on the short-range solver (the 5% of the code that change from platform to platform) and kept the long-range force solver unchanged. Given that the starting version was already optimized for the BG/Q, another many-core system, the port to Theta went very smoothly.

5.2 Halo Finder in HACC CosmoTools: Achieving Portability via Thrust

Halo finding is a key analysis task for both gravity-only and hydrodynamics simulations. While in gravity-only simulations this task can be carried out in post-processing, for the hydrodynamic simulation an in-situ implementation of the halo finder is essential. Sub-grid models such as AGN feedback rely on knowledge of the halo positions and masses during the evolution of the simulation. Given the diverse architectures HACC is running on already and CRK-HACC will be in the future, a fully portable approach to analysis tools is very desirable to ensure best performance for in-situ analysis. We have found that for the HACC halofinder the Thrust library provides an ideal platform to realize portability combined with excellent performance. In the following, we show comparisons of time-to-solution for the halo finder using a realistic test problem on different architectures. We would like to emphasize that the code itself was not changed at all for this test when used on the different machines. The Thrust library is used for the center finding algorithm, Figure 14 shows the comparison for the different machines for this operation. All other timings are given for completeness.

We compare halo finding performance on Mira (BG/Q), Cooley (GPU-enhanced with NVIDIA K80s), and Theta (KNL); details of these systems are given below, followed by tables of the max, min, and average times for each step of the analysis per rank, from a single run of the code on each

Table 6: Maximum runtimes in seconds for halo finding stages

Stage	Mira	Cooley/Thrust	Theta/Thrust	Theta/No-thrust
fof	220.690	19.418	49.468	49.483
particle-i/o	12.694	2.632	4.005	4.204
fof-properties	1.601	0.364	0.581	0.584
center-finder	1794.900	131.450	73.815	461.390
halo-catalog	6.737	0.926	0.920	0.748
core-finder	35.370	4.303	9.872	9.051
sod-finding	9.519	0.942	2.353	2.349
sod-properties	3.097	0.165	0.896	0.641
sod-bin-profile	1.922	0.417	0.454	0.471
subfind	10465.000	920.910	2864.100	2863.700
sub-halo-properties	3.426	1.555	2.138	2.184
LANLHaloFinder	12555.000	1083.100	3008.700	3394.900

Table 7: Average runtimes in seconds for halo finding stages

Stage	Mira	Cooley/Thrust	Theta/Thrust	Theta/No-thrust
fof	220.690	19.418	49.457	49.472
particle-i/o	12.694	2.631	3.990	4.190
fof-properties	0.616	0.154	0.152	0.152
center-finder	266.560	54.093	2.629	41.576
halo-catalog	6.737	0.926	0.919	0.748
core-finder	35.369	4.303	9.872	9.051
sod-finding	4.597	0.404	0.767	0.769
sod-properties	3.096	0.164	0.896	0.641
sod-bin-profile	1.922	0.416	0.454	0.471
subfind	2583.500	233.370	560.480	560.540
sub-halo-properties	3.425	1.555	2.137	2.183

system (Tables 6-8). Cooley far outperforms Mira due primarily to its more powerful processors; the Thrust/GPU-based center-finder yields more than a 10× speedup over Mira. The Thrust-based center-finder also utilizes the Theta cores well, yielding more than 1.7× speedup over Cooley, demonstrating the performance portability of this phase of the computation. Theta results are also given without the use of Thrust for comparison, in which case all phases are similar with the exception of center-finding. Figure 14 shows a summary of the timings for the center finder, by far the most expensive portion of the halo finder.

Mira
Node configuration: IBM BQC (16 core)
Performance: Single-node, double-precision 204.8 GFlops
Run configuration: 256 nodes, 1 rank per node

Cooley
Node configuration: 2 × Intel Haswell E5-2620 v3 (2 × 6 core); 1 × NVIDIA K80 GPU
Performance: Single-node, double-precision 1.87 TFlops (K80, no boost)
Run configuration: 64 nodes, 4 ranks per node

Theta
Node configuration: Intel Xeon Phi KNL (64 cores)

Table 8: Minimum runtimes in seconds for halo finding stages

Stage	Mira	Cooley/Thrust	Theta/Thrust	Theta/No-thrust
fof	220.690	19.417	49.456	49.471
particle-i/o	12.693	2.631	3.979	4.179
fof-properties	0.157	0.055	0.032	0.032
center-finder	18.388	23.768	0.165	1.345
halo-catalog	6.736	0.926	0.919	0.748
core-finder	35.369	4.303	9.871	9.050
sod-finding	2.043	0.138	0.269	0.268
sod-properties	3.096	0.164	0.896	0.641
sod-bin-profile	1.922	0.416	0.454	0.471
subfind	203.910	19.616	11.402	11.412
sub-halo-properties	3.425	1.555	2.137	2.183

Performance: Single-node, double-precision 2.6 TFlops
Run configuration: 64 nodes, 8 ranks per node, 16 threads per rank

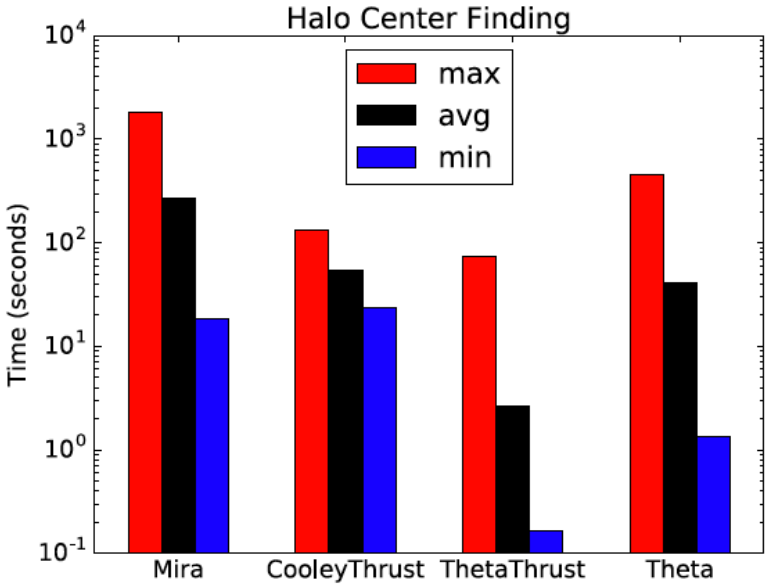

Figure 14: Halo Center Finding Timings
Summary of the halo center finding timings for HACC on different architectures with and without Thrust. On Theta, the use of Thrust delivers a large benefit, on Cooley, the Thrust implementation allows the GPU utilization for this calculation. The timings shown are for the maximum, minimum, and average time spent per node.

6 Conclusions

During this ESP we have successfully ported HACC to Theta and demonstrated very good performance on the KNL architecture. We have developed a completely new capability for HACC and implemented a hydrodynamic solver and a first set of subgrid models. CRK-HACC will play an

important role in our group's future cosmology program. Developing a cosmological hydrodynamics code that can scale up to the largest machines available is a very important step for the cosmology community and will have a large impact for upcoming cosmological surveys such as LSST, WFIRST, and DESI. CRK-HACC on the KNL architecture will be able to successfully fill this need.

Acknowledgements

This research was in part supported by the Exascale Computing Project (ECP), Project Number: 17-SC-20-SC, a collaborative effort of two DOE organizations—the Office of Science and the National Nuclear Security Administration—responsible for the planning and preparation of a capable exascale ecosystem—including software, applications, hardware, advanced system engineering, and early testbed platforms—to support the nation's exascale computing imperative.

References

[1] N. Battaglia, J. R. Bond, C. Pfrommer, J. L. Sievers, and D. Sijacki. Simulations of the Sunyaev-Zel'dovich Power Spectrum with Active Galactic Nucleus Feedback. *Astrophys. J.*, 725:91–99, December 2010.

[2] C. S. Frenk, S. D. M. White, P. Bode, J. R. Bond, G. L. Bryan, R. Cen, H. M. P. Couchman, A. E. Evrard, N. Gnedin, A. Jenkins, A. M. Khokhlov, A. Klypin, J. F. Navarro, M. L. Norman, J. P. Ostriker, J. M. Owen, F. R. Pearce, U.-L. Pen, M. Steinmetz, P. A. Thomas, J. V. Villumsen, J. W. Wadsley, M. S. Warren, G. Xu, and G. Yepes. The Santa Barbara Cluster Comparison Project: A Comparison of Cosmological Hydrodynamics Solutions. *Astrophys. J.*, 525:554–582, November 1999.

[3] F. Haardt and P. Madau. Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. *Astrophys. J.*, 461:20, April 1996.

[4] F. Haardt and P. Madau. Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background. *Astrophys. J.*, 746:125, February 2012.

[5] S. Habib, V. Morozov, H. Finkel, A. Pope, K. Heitmann, K. Kumaran, T. Peterka, J. Insley, D. Daniel, P. Fasel, N. Frontiere, and Z. Lukic. The Universe at Extreme Scale: Multi-Petaflop Sky Simulation on the BG/Q. *ArXiv e-prints*, November 2012.

[6] S. Habib, V. Morozov, N. Frontiere, H. Finkel, A. Pope, and K. Heitmann. HACC: Extreme Scaling and Performance Across Diverse Architectures. In *SC '13 Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis, Article #6*, November 2013.

[7] S. Habib, A. Pope, H. Finkel, N. Frontiere, K. Heitmann, D. Daniel, P. Fasel, V. Morozov, G. Zagaris, T. Peterka, V. Vishwanath, Z. Lukić, S. Sehrish, and W.-k. Liao. HACC: Simulating sky surveys on state-of-the-art supercomputing architectures. *New Astronomy*, 42:49–65, January 2016.

[8] S. Habib, A. Pope, Z. Lukić, D. Daniel, P. Fasel, N. Desai, K. Heitmann, C.-H. Hsu, L. Ankeny, G. Mark, S. Bhattacharya, and J. Ahrens. Hybrid petacomputing meets cosmology: The

Roadrunner Universe project. In *Journal of Physics Conference Series*, volume 180 of *Journal of Physics Conference Series*, page 012019, July 2009.

[9] K. Heitmann, N. Frontiere, C. Sewell, S. Habib, A. Pope, H. Finkel, S. Rizzi, J. Insley, and S. Bhattacharya. The Q Continuum Simulation: Harnessing the Power of GPU Accelerated Supercomputers. *Astrophys. J.S.*, 219:34, August 2015.

[10] K. Heitmann, P. M. Ricker, M. S. Warren, and S. Habib. Robustness of Cosmological Simulations. I. Large-Scale Structure. *Astrophys. J.S.*, 160:28–58, September 2005.

[11] N. Katz, D. H. Weinberg, L. Hernquist, and J. Miralda-Escude. Damped Lyman-Alpha and Lyman Limit Absorbers in the Cold Dark Matter Model. *Astrophys. J.L.*, 457:L57, February 1996.

[12] V. Springel and L. Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *MNRAS*, 339:289–311, February 2003.

[13] J. Woodring, K. Heitmann, J. Ahrens, P. Fasel, C.-H. Hsu, S. Habib, and A. Pope. Analyzing and Visualizing Cosmological Simulations with ParaView. *Astrophys. J.S.*, 195:11, July 2011.

Argonne Leadership Computing Facility
Argonne National Laboratory
9700 South Cass Avenue, Bldg. #240
Argonne, IL 60439

www.anl.gov

Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC